

From Legacy to the Cloud: Novo Nordisk's Journey towards a Modern Statistical Computing Environment

Marianne Caramés, Novo Nordisk A/S, Copenhagen, Denmark

Andrew York, Novo Nordisk A/S, London, UK

ABSTRACT

Novo Nordisk A/S has embarked on a journey to replace their legacy UNIX-based SCE with a more modern, multi-language, cloud-based solution. We will present our journey towards a new SCE and the challenges we have encountered along the way. Taking points in our current setup and how our thoughts and ideas changed over time. As a result, our ambition of having an SCE covering our current needs and requirements changed towards getting a modular and modern solution, which could enable fast adoption of new tools, automation, and innovation. To this end, Novo Nordisk scored the 8 different solutions and vendors based on the Request for Quotation material received. We ended up selecting two vendors to participate in a staged proof of concept.

INTRODUCTION

Novo Nordisk A/S has for a couple of years searched for a replacement of the on-premise HP Unix system used in the production of deliverables for clinical study reports and submissions. The search has brought us to cloud environments, CDR solutions and analysis of requirements for the future use of data and production of deliverables.

ANALYSIS AND PREPARATION

REQUEST FOR QUOTATION (RFQ)

In the second half of 2022 Novo Nordisk A/S decided to do an RfQ with possible suppliers of SCE systems. We engaged with 9 different suppliers. The Novo Nordisk team prepared RfQ templates material to be filled out including our requirement for the new SCE. The suppliers received the RfQ material, and 8 of them decided to respond to the RfQ. The received RfQ material was evaluated by the team using a huge evaluation sheet investigating supplier functionality such as Process Management, Automation, Production of deliverables for submission, System design, User Experience, Competences and capabilities, Technical solution, Supplier capacity and Price. The suppliers demoed their suggested solution to the Novo Nordisk team. The solutions presented were very different. Some were plug&play SCE solutions, and some were building block based in nature. The plug&play solutions could bring us to a fast implementation, but were they flexible enough for us to keep on developing and exchanging modules?

PROOF OF CONCEPT (POC)

An evaluation meeting was held to decide on the supplier(s) to engage in a PoC. During the meeting and discussions, the opinions changed towards selection of a supplier with an open and flexible solution, even though we could see that some gaps would need to be mitigated in order to have the SCE platform, we needed. We decided to do a staged PoC starting with the supplier with the open and flexible solution, Domino Data Lab, and if we evaluated the PoC negative, then do a second PoC with the second-best solution.

Several extra users from the coming users joined for the PoC, and a subset of the requirements were selected to test. We decided to use the testing environment setup by Domino for the PoC and add some test data. This approach also resulted in the testers not being able to test some of the requirement like interfaces to sourcing systems and external applications to use. A huge excel sheet was developed, so each requirement could be tested by the users, and at the same time we did a gap analysis of the solution. The testing of the PoC was finalized without any showstoppers. The gap analysis showed that functionality within the following areas needed to be developed:

- Access model/control – we want to have a template/script developed to ensure consistency of folders and accesses setup for each study
- Automation/data lineage – this is an area, where the solution did not offer any functionality, so we have engaged with a sub-contractor to develop a setup, where downstream programs are executed dependent on changes to data, metadata or programs

- Audit trail – the solution has a huge audit trail, but for inspection/audit purposes and operation & maintenance activities we will like to have some reports developed allowing to filter on specific studies, accesses or other areas
- Package manager for R/Python – we are already using a package manager for our R/Python development area and will like to reuse this and/or the images created for R/Python packages

DEVELOPMENT

The contract was signed end-March 2023. We have re-branded the project to AMACE: Automated Multi-lingual Analytic Computing Environment and restructured the NN team into 3 teams, each working on a separate area:

- **DataDepot:** storage (CDR) and infrastructure
- **DataCrunch:** compute (Domino), tools, automation, workflows and gaps
- **Wizard:** end-user applications

We have now on-boarded the competencies needed and the installation, configuration and development have started.

DESIGN DECISIONS

During the creation of requirements for the solution the following design decisions were identified:

- Use of NN AWS public cloud – to streamline with other solutions in Novo Nordisk A/S we want to ensure a seamless transfer and publication of our results and data
- Language agnostic (SAS, R, Python, etc.) – many new employees come with experience in other programming languages, so we want our solution to be language agnostic offering 3 programming languages as a start but with the possibility to add other when needed
- Simple but granular access model/control (able to work with our Novo Nordisk access control system and Azure Active Directory) – a simpler access model than the existing one in the current system is developed
- Support for GxP compliant submissions (understandable audit trail) – reports on top of the audit trail is developed to make sure that it was readable and understandable and can be used in audits/inspections, if needed
- Able to use our GIT repos in Novo Nordisk A/S – several of our clinical project are already using GIT repositories for their programs, so we need to ensure that this work could continue in the new solution
- Efficient access to tools and data (system tools, access to sourcing data/metadata) – a longer list of system tools has been defined and seamless access to our sourcing systems will be set up using APIs when possible
- Sandbox environment (supporting innovation and experimentation) – in the new solution it will be easy to set up and test out new tools, programming languages, versions or packages
- Modular approach to add upcoming technologies – modularity is key to be able to continuously exchanging part of the solution when something better is discovered
- Support of automation and data lineage – in order to effectively react on changes to data, metadata or programs, it is a must to be able to identify the changes to incoming data/metadata and programs and only execute the downstream impacted programs

This resulted in the following vision: **Deliver an automated, scalable, multilingual setup to enable fast analysis of clinical data, facilitate collaboration and enable reusability for end-users**

CHALLENGES

A decision was made to setup our solution outside our corporate environment in order to have a better network performance, but this decision has also resulted in a more cumbersome setup of our interfaces to other systems.

To align with other solutions we decided to use our internal AWS DataHub solution and create our Clinical Data Repository (CDR) in the offered S3 buckets. This meant that we had a limitation on the number of S3 buckets (1000 per account), due to a restriction set up in the DataHub solution. Also, we did not have access to a FS drive, which made it difficult to navigate and browse the files in the buckets.

Roles/access to be set up in Domino and the DataHub solution had to be aligned.

Credential propagation from the DataHub solution to Domino was difficult to setup. It was found that S3 buckets/AWS did not understand the AD groups, so an O2 platform was added to have the SAML tokens created for the access.

Based on the challenges we were forced to ask the question: Is the system we are building scalable and granular enough?

NEXT STEP

To answer the question it has been decided to test out a solution with Azure, DataBricks, Domino and SAS Viya in a PoC. We will set up and test access, integration to Novo Nordisk access control system/AD, programming from both Domino and SAS Viya and usability of navigating and browsing data.

CONCLUSION

The conclusion of the PoC based on Azure is still to be made and a strategy for cloud environment will follow. We hope to be wiser when we see you in November.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Marianne Caramés
Novo Nordisk A/S
Vandtårnsvej 114
DK-2860 Søborg
Work Phone: +45 30 79 12 11
Email: maca@novonordisk.com
Web: www.novonordisk.com